

# An Evaluation of Critical Thinking in Competency-Based and Traditional Online Learning Environments

Matthew Mayeshiba  
*University of Wisconsin-Extension*

Kay R. Jansen and Lisa Mihlbauer  
*University of Wisconsin-Milwaukee*

## Abstract

Nonterm, direct assessment competency-based education (CBE) represents a significant reimagining of the structure of higher education. By regulating students' progress through the program based on their mastery of tightly defined competencies rather than on the time spent learning them, this learning environment affords students far greater flexibility than traditional programs. This focus on defined competencies has led to concerns that students in these types of programs may not demonstrate higher level skills, such as critical thinking, at levels comparable to those enrolled in more traditional programs. This study evaluated 39 students' demonstration of critical thinking in two assessments administered in parallel versions of one course: one offered through the nonterm, direct assessment CBE University of Wisconsin Flexible Option, and the other offered through a traditional online program. For this study, each of the 78 assessments was scored using the critical thinking rubric from the Valid Assessment of Learning in Undergraduate Education (VALUE) project. We found that students from the CBE version of the course received significantly higher ( $p = .0013$ ) overall scores than the students in the traditional online version of the course. While further research is required to refine these methods and ensure the generalizability of these results, they do not support concerns about students' abilities in this learning environment.

*Keywords:* AAC&U, competency-based education, direct assessment, disruptive innovation, critical thinking, learning outcomes, non-term, nonterm, UW Flexible Option, VALUE rubric

Mayeshiba, M., Jansen, K.R., & Mihlbauer, L. (2018). An evaluation of critical thinking in competency-based and traditional online learning environments. *Online Learning, 22*(2), 77-89. doi: 10.24059/olj.v22i2.1365

## **An Evaluation of Critical Thinking in Competency-Based and Traditional Online Learning Environments**

Competency-based education (CBE) has been a focal point for recent efforts to offer pathways through postsecondary education that are more responsive to the needs of nontraditional students, who require more flexibility than can be offered in traditional programs (Eduventures, 2014). Nonterm, direct assessment CBE programs, such as the University of Wisconsin (UW) Flexible Option, give students a wider array of possible start dates and allow students to progress through the material as soon as they demonstrate mastery. This offers students many more opportunities to tailor their learning experience so they can leverage previous experiences and accommodate outside obligations.

This paper outlines a still preliminary effort to document the higher level competencies of students in a direct assessment CBE program. To do this, the authors of this study examined student critical thinking in two parallel versions of one course: one version through the nonterm, direct assessment UW Flexible Option, and another version offered through a parallel traditional online program. This study leverages the similarities between these two courses to investigate the following research question:

- Do students in the UW Flexible Option demonstrate critical thinking at levels similar to those demonstrated by students enrolled in a comparable traditional online environment?

### **Review of Related Literature**

While the concept is not new, interest in CBE programs has increased in recent years as institutions of higher learning have sought scalable methods of becoming more accessible to nontraditional students (Nodine, 2016). With the emphasis on demonstrated mastery rather than measured seat time, CBE programs have implemented different models to ensure students have greater flexibility in structuring their studies. These models range from maintaining a close resemblance to traditional academic calendars, through various subscription models, to allowing students to move entirely at their own pace (Kelchen, 2015).

This focus on demonstrated mastery, however, has raised concerns about the role of higher level learning objectives in these programs. Ward (2016) raised concerns that CBE programs inadequately focus on broad-based learning objectives that are difficult to measure, even though there is evidence that these learning objectives are in high demand among employers (Hart Research Associates, 2015) and of great social value. If such skills are not adequately incorporated into the learning curriculum, the degrees awarded by such programs would fundamentally be of less value, leading to further stratification of higher education into those students who receive a “good enough” education and those who receive a quality one (Ward, 2016).

In response to this, CBE advocates have identified a number of best practices to ensure the integrity of academic offerings, including robust engagement with multiple stakeholders (CAEL, 2014), explicit mapping of competencies and learning experiences (Johnstone & Soares, 2014), and robust efforts to engage with students throughout the learning process (Gruppen, 2016). Additionally, Krause, Dias, and Schedler (2016) have tested a framework to codify good course design features in CBE. Central Washington University established a rubric to support their CBE FLEX-IT program to evaluate course design elements and found correlations with student assessment scores (2017).

Despite the above efforts, very little empirical research has been conducted to quantify the higher level competencies demonstrated by CBE students. In fact, a review of the literature reveals only one attempt to measure general education outcomes among a “small sample” of students at the College for America, a subsidiary of Southern New Hampshire University and one of the first institutions in the United States to provide postsecondary degrees through direct assessment (Fain, 2015). This effort, reported only in the popular press, used the Proficiency Profile from the Educational Testing Service to assess student skills in critical thinking, reading, writing, mathematics, humanities, social sciences, and natural sciences. This effort showed that the CBE students outperformed the benchmark group in all areas except mathematics. Despite this apparent success, the lack of precise information on the sample size and population, as well as the study’s lack of peer review, limit the usefulness of the effort.

## Methods

### **UW Flexible Option: A Nonterm, Direct Assessment CBE Program**

The UW Flexible Option was established in January 2014 as an interinstitutional partnership led by UW-Extension on behalf of UW System Administration and in collaboration with various UW campuses. As one of the first adopters in what Nodine (2016) called the third generation of CBE providers, the UW Flexible Option distinguished itself by offering postsecondary degrees in a nonterm, direct assessment learning environment. Unlike students in traditional learning environments, UW Flexible Option students move through the program at a rate based on their demonstrated mastery of the material, rather than on the time they have spent studying it. As a result, students in this program enroll in a series of three-month subscription periods that begin at the start of every calendar month. To facilitate additional flexibility, students have no deadlines by which they need to complete their work, and students are allowed to carry uncompleted coursework from one subscription to another without penalty or special considerations using an “In Progress” grade. These factors allow students to move more quickly through material they already understand, or more slowly when their learning or outside commitments demand it.

These flexibilities necessitate significant changes for the teaching and learning experiences in this program. First, the ease with which students can stop out and reenter the program to accommodate their outside obligations means that students do not move through the program with any consistent cohort of other students. Additionally, because there are no set deadlines for submitting assigned work, even students who are enrolled in the same course at the same time may be engaging with very different parts of the curriculum at any given moment. As a result, the established mechanisms for interstudent interaction found in traditional online programs, such as discussion boards, are not applicable to the UW Flexible Option.

This reality changes a number of aspects of both the teaching and learning experience. For instance, faculty members must be much more careful and explicit in their curation of learning materials and assessments. Additionally, it becomes much more important that students have regular interaction with a broad student support network, including faculty, tutors, academic coaches, and others.

## The Critical Thinking VALUE Rubric

This project defined and operationalized the term *critical thinking* using the Valid Assessment of Learning in Undergraduate Education (VALUE) rubric sponsored by the Association of American Colleges and Universities (2016). Assembled between 2007 and 2009 by teams of faculty and other higher education professionals from more than 100 institutions of higher education, this set of 16 rubrics provides a framework for operationalizing student demonstration of a variety of metacognitive skills. These rubrics have been widely distributed within higher education, having been accessed by more than 42,000 individuals from more than 4,200 unique institutions as of December 2015 (AAC&U, 2016).

The VALUE rubric employed for this study defines critical thinking as “a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.” The rubric breaks this larger concept into five distinct dimensions: explanation of issues, evidence (selecting and using information to investigate a point of view or conclusion), influence of context and assumptions, student’s position (perspective, thesis/hypothesis), and conclusions and related outcomes (implications and consequences). Finally, each dimension is broken into five separate performance levels, scored from 0 (*not present*) to 4 (*capstone*), with language describing the depth of skill demonstrated at each level.

The VALUE rubrics are widely used throughout higher education as a tool for measuring student demonstration of metacognitive skills to facilitate a better understanding of what students know and can do. The AAC&U website documents practices at a wide variety of institutions that have used these rubrics to assess student work from within individual courses, at the program level, and institution-wide to assess student demonstration of broader learning objectives. Additionally, the Multi-State Collaborative to Advance Learning Outcomes Assessment (MSC) is an effort led by the State Higher Education Executive Officers Association (2016) that is currently underway to reliably and robustly measure student demonstration of metacognitive skills across 12 states and 88 two- and four-year campuses.

## Course Structure and Assessment Context

The UW Flexible Option and traditional online version of the course that this project examined were hosted by the same University of Wisconsin institution, relied on the same curriculum, and used assessments that had been specifically tuned to incorporate the same assignment prompts and grading rubrics. Nevertheless, significant differences did remain between the two courses. First, instructors between the two versions of the course were not the same. For this project all UW Flexible Option students were evaluated by one instructor, while the traditional online students were split among three different instructors. Additionally, the traditional online course mandated participation in a variety of activities and discussions separate from the scored assessments, while the asynchronous nature of the UW Flexible Option meant that opportunities for this sort of interstudent interaction were not present in that version of the course.

Additionally, the two assessments examined here also were situated within very different contexts based on the expectations of their learning environments. Students in the traditional online course were presented with a series of deadlines for submitting their assessments. These deadlines fell roughly six weeks apart with several activities and mandated feedback occurring between the two dates. These deadlines were not incorporated into the UW Flexible Option version of the course, and students were free to submit either of their assessments at any time during their

subscription. Some students allowed significant time to pass between submitting these two assessments, while others submitted the two assessments at nearly the same time. Still other students submitted the two assessments out of order. This behavior is consistent with the flexibilities built into the nonterm, direct assessment design of the UW Flexible Option.

Finally, it is important to note that in these two versions of the course the two assessments were presented in opposite orders. For the traditional online students, Assessment A was due roughly halfway through the semester, while the deadline for Assessment B fell just before the end of the term. In the UW Flexible Option, however, Assessment A was presented to the students as Assessment #2, while Assessment B was referred to as Assessment #1 in course materials. As a result, the vast majority of UW Flexible Option students submitted Assessment B before submitting Assessment A. This complicates the interpretation of the findings. Scores might be expected to increase as students move through the course because of a variety of factors related to student learning, including the incorporation of instructor feedback and deeper exposure to the material. Because these two assessments were presented in opposite orders, it can become more difficult to understand the role of the different learning environment as opposed to the role of these learning effects. For the analysis presented below, however, our results show that traditional online students did not outperform the students enrolled in the UW Flexible Option version of the course on either assessment. Even on Assessment B, where these learning effects should have been largest for the traditional online students and smallest for the UW Flexible Option students, average scores for students in the traditional online version of the course were not higher than those of students enrolled in the UW Flexible Option. Therefore, we believe that this effect does not undermine the essential finding of the paper.

### **Scoring Process**

For this project, two senior faculty from the course's department scored student work samples from 39 students enrolled in parallel versions of a single course. Of these students, 15 were enrolled in a version of the course offered through the UW Flexible Option, while the remaining 24 were enrolled in a course offered through a traditional online degree program. For each student, faculty scored two assessments, both of which were papers with a maximum length of 10 double-spaced pages and submitted as part of the students' course grade. Both faculty scorers were familiar with the course content, and in one case had taught the course during previous terms. Neither, however, had been involved in teaching either version of the course during the project period.

Once students completed their coursework, the lead analyst randomly identified a sample of traditional online students for inclusion in the study. Because the number of students enrolled in the UW Flexible Option version of the course was relatively small, all student work from that version of the course was included. The analyst then created de-identified copies of each assessment that would be scored by converting the submitted work samples into a unified format, removing personally identifying information, such as names, ages, or places of work, as well as removing information identifying the program of study, such as the course number, name of the instructor, or the program name. Assessments were then assigned a random artifact identifier and presented for scoring.

Prior to scoring work included in the study, the faculty scorers participated in a calibration session with a nationally recognized expert in the VALUE rubrics. This process involved a guided scoring session in which the scorers evaluated two assessments written by traditional online

students whose work was not included as part of the randomly drawn sample. After scoring each assessment, the scorers and calibration leader discussed their scores and mutually agreed upon how to define and operationalize the terms of the rubric.

For the scoring itself, the faculty scorers read each of the de-identified assessments and assigned a whole number score from 0 to 4 for each dimension of the VALUE rubric. Due to the number of assessments, this process took several weeks, with scorers occasionally comparing scores on completed assessments to ensure continued calibration. Additionally, once scoring was complete, the overall results were checked, and cases where the two scorers differed on one dimension by more than one point were identified. These cases then were referred to the scorers for review, and scorers were given the opportunity to revise the scores to ensure they represented a consistent understanding of the rubric among the two scorers. Of the 390 dimensions scored on the 78 separate assessments examined, 19 such cases were identified in 10 separate assessments. Once this process was complete, the two scores submitted for each dimension were averaged to arrive at a final score for each dimension of the rubric.

To measure the reliability of the scoring process, this analysis applied Cohen’s kappa statistic with linear weighting to the results recorded both before and after the reconciliation process. In this case, the kappa statistic measures the degree to which the two faculty members agreed on the score assigned to each dimension of the rubric relative to the odds that the scores would have agreed by chance (Cohen, 1960). Further, because the scale for each dimension was ordinal, a linear weighting procedure was applied that gives partial credit for answers that were close (Cohen, 1968). This scale ranges from -1 to 1 with 1 indicating perfect agreement, -1 indicating perfect disagreement, and 0 indicating agreement equal to what would have been demonstrated if the scores were assigned randomly. For this statistic, scores in excess of .20 are typically considered fair agreement, scores in excess of .40 are typically considered to be in moderate agreement, and scores in excess of .60 are typically considered to be in substantial agreement (Viera & Garrett, 2005). Kappa statistics for both reconciled and unreconciled scores are presented in Table 1 (Lowry, 2016). These statistics indicate that the reconciled scores achieved a linear weighted agreement of .4084 ( $\pm .0609$ ), indicating moderate agreement between the two scorers.

Table 1.

*Unreconciled and Reconciled Kappa Statistics*

	Kappa statistic	Std. error	95% confidence interval	
			Lower bound	Upper bound
<i>Unreconciled scores</i>				
Unweighted	.2375	.0375	.1641	.3109
Linear weighted	.3607	.0337	.2946	.4268
<i>Reconciled scores</i>				
Unweighted	.2382	.0374	.1649	.3115
Linear weighted	.4084	.0310	.3475	.4693

*Note.* Unreconciled scores are the scores awarded before dissimilar results were reconciled through additional discussion between scorers, while reconciled scores are scores awarded after this process. Unweighted kappa statistics are those that do not award partial credit for scores that were close, while linear weighted kappa statistics awarded half credit for scores that differed by only one point.

### Data Analysis

Data analysis for this project examined average reconciled scores for each assessment individually and both assessments overall for each student whose work was scored. Combining these variables into a set of average scores allows for a clearer aggregate look at student performance between these two delivery modalities. At the same time, an analysis of correlations among the variables involved demonstrates that the dimension-level scores are reliably related and that using these aggregate measures does not significantly influence the result. A full correlation matrix for all dimensions of both assessments is presented in Table 2. Dimensions for Assessment A are represented as variables 1 through 5 on this table. These variables demonstrate statistically significant correlations among the final scores awarded for each dimension of the rubric. Furthermore, the standardized Cronbach’s alpha coefficient of .879 further supports the utility of a combined measure. Dimensions for Assessment B are represented as variables 6 through 10 on this table. These also demonstrate the high degree of correlation among the dimension-level scores that result in a standardized Cronbach’s alpha coefficient of .942. Finally, the correlation matrix for all 10 dimensions of both assessments further supports the combination of these variables and presents a standardized Cronbach’s alpha coefficient of .890.

Table 2.

*Correlation Matrix for All Dimension-Level Variables*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Explanation (Assessment A)	1.000									
2. Evidence (Assessment A)	.484 (.002)	1.000								
3. Context (Assessment A)	.340 (.034)	.567 (<.001)	1.000							
4. Position (Assessment A)	.481 (.002)	.614 (<.001)	.819 (<.001)	1.000						
5. Conclusions (Assessment A)	.400 (.012)	.595 (<.001)	.764 (<.001)	.849 (<.001)	1.000					
6. Explanation (Assessment B)	.496 (.001)	.325 (.043)	.284 (.080)	.220 (.178)	.124 (.451)	1.000				
7. Evidence (Assessment B)	.354 (.027)	.320 (.047)	.138 (.401)	.167 (.309)	-.048 (.771)	.630 (<.001)	1.000			
8. Context (Assessment B)	.313 (.052)	.344 (.032)	.206 (.209)	.202 (.218)	.039 (.815)	.662 (<.001)	.834 (<.001)	1.000		
9. Position (Assessment B)	.435 (.006)	.456 (.004)	.314 (.051)	.290 (.073)	.092 (.579)	.676 (<.001)	.871 (<.001)	.857 (<.001)	1.000	
10. Conclusions (Assessment B)	.338 (.035)	.481 (.002)	.290 (.074)	.298 (.066)	.094 (.569)	.628 (<.001)	.842 (<.001)	.797 (<.001)	.866 (<.001)	1.000

*Note.* This table presents correlation coefficients (and confidence intervals) for all dimension-level variables for Assessment A and Assessment B. Variables 1 through 5 were included in the Assessment A average score and have a standardized Cronbach’s alpha coefficient of .879. Variables 6 through 10 were included in the Assessment B average score and have a standardized Cronbach’s alpha coefficient of .942. Variables 1 through 10 were included in the total average score and have a standardized Cronbach’s alpha coefficient of .890.

Except for student age, data describing the traditional online student population were unavailable for this analysis. To examine this variable’s importance, this analysis included a linear regression of average total score, the student’s age in years, and a quadratic age term. The results of this regression are presented in Table 3. These results indicate that both age ( $p = .0101$ ) and its quadratic term ( $p = .0110$ ) are significant predictors of a student’s score, with a maximum predicted score among students who are 37.94 years old. For this reason, both terms will be included in further analyses.

Table 3.

<i>Linear Regression of Age and Overall Score</i>				
Variable	Estimate	Standard error	<i>t</i> -value	<i>p</i> -value
Intercept	-2.5585	1.6009	-1.60	.1188
Age	0.2399	0.0884	2.71	.0101 *
Age squared	-0.0032	0.0012	-2.68	.0110 *

*Note.* Results in this table are over 39 students and have an adjusted *R*-squared value of 12.4%.  
\* $p < .05$ .

Using the three average score variables described above, this paper’s main analysis examined student performance in both assessments combined and then within each assessment. Therefore, the analysis relied on a set of three linear regressions of the following form:

$$Average\ Score_i = \beta_0 + \beta_1(Age_i) + \beta_2(Age\ Squared_i) + \beta_3(Program_i) + \varepsilon_i$$

In this equation, *Average Score<sub>i</sub>* is the average score of student *i* on either the first assessment, second assessment, or across both assessments. *Age<sub>i</sub>* and *Age Squared<sub>i</sub>* are the student’s age and its quadratic term. *Program<sub>i</sub>* is a dummy variable indicating the version of the course in which the student enrolled, where enrollment in the UW Flexible Option is coded as 1 and enrollment in the traditional online version of the course is coded as 0. The results of this analysis are described in the section below.

Finally, this analysis used a paired *t*-test to examine changes in each student’s scores between the two assessments. The paired *t*-test is used to compare changes in means when each subject in a study is measured at two points in time. Because it measures differences in scores for each student, this test provides an indication of whether each student’s scores changed from Assessment A to Assessment B.

## Results

The first analysis conducted here investigated the role of the course version on the student’s overall average score. The results of this regression are detailed in Table 4 and indicated that, overall, students in the UW Flexible Option received higher scores than students in traditional online versions of the course. This linear regression demonstrated that students enrolled in the UW Flexible Option version of the course scored 0.44 points higher on average in each dimension across both assessments. This difference was statistically significant ( $p = .0013$ ). Additionally, both age ( $p = .0045$ ) and its quadratic term ( $p = .0050$ ) were also statistically significant, indicating an approximate age of maximum score at 39.10 years.

Table 4.

*Linear Regression of Age and Course Version Against Overall Average Score*

Variable	Estimate	Standard error	t-value	p-value
Intercept	-2.6445	1.39926	-1.89	.0671
Age	0.2347	0.07725	3.04	.0045 **
Age squared	-0.0031	0.00103	-3.68	.0050 **
UW Flexible Option	0.4499	0.12911	3.48	.0013 **

Note. Results in this table are over 15 UW Flexible Option students and 24 traditional online students and have an adjusted R-squared value of 33.1%.

\*\* $p < .01$ .

To further investigate these results, this analysis included two more linear regressions investigating each student's average score in each of the two assessments. For Assessment A, while UW Flexible Option students retained an average score that was 0.15 points per dimension higher than traditional online students, this difference was not statistically significant ( $p = .3146$ ). On the other hand, age ( $p = .0006$ ) and its quadratic term ( $p = .0007$ ) were both statistically significant, with an approximate age of maximum score at 38.34 years. These results are further illustrated in Table 5. For Assessment B, the statistical significance of these results was reversed, with UW Flexible Option students receiving scores that were on average 0.75 points per dimension higher ( $p = .0002$ ) but with statistically insignificant effects for age ( $p = .1920$ ) and its quadratic term ( $p = .1904$ ). These results are further illustrated in Table 6.

Table 5.

*Linear Regression of Age and Course Version Against Assessment A Average Score*

Variable	Estimate	Standard error	t-value	p-value
Intercept	-4.1101	1.54554	-2.66	.0117 *
Age	0.3236	0.08532	3.79	.0006 **
Age squared	-0.0042	0.00114	-3.71	.0007 **
UW Flexible Option	0.1455	0.14261	1.02	.3146

Note. Results in this table over 15 UW Flexible Option students and 24 traditional online students and have an adjusted R-squared value of 25.2%.

\*  $p < .05$ . \*\*  $p < .01$ .

Table 6.

*Linear Regression of Age and Course Version Against Assessment B Average Score*

Variable	Estimate	Standard error	t-value	p-value
Intercept	-1.1788	1.98485	-0.59	.5564
Age	0.1458	0.10957	1.33	.1920
Age squared	-0.0020	0.00146	-1.34	.1904
UW Flexible Option	0.7542	0.18315	4.12	.0002 **

Note. Results in this table are over 15 UW Flexible Option students and 24 traditional online students and have an adjusted R-squared value of 29.6%.

\*\* $p < .01$ .

To further investigate this difference in scores, the analysis continued with a paired *t*-test to evaluate the difference between Assessment A and Assessment B scores for each student based on the version of the course they enrolled in. These results are illustrated in Table 7. The results of this test showed that UW Flexible Option students scored better on Assessment A by 0.2333 points per dimension and that this difference was statistically significant at  $\alpha = .05$  ( $p = .0440$ ). Additionally, this test demonstrated that the traditional online students scored better on Assessment B by 0.3708 points per dimension and that this difference was also statistically significant at  $\alpha = .05$  ( $p = .0186$ ). In both cases, students scored better on the second assessment that they were presented.

Table 7.

*Paired t-Tests on Changes in Score From Assessment A to Assessment B*

Course version	Difference in mean	Standard error	<i>t</i> -value	<i>p</i> -value
Traditional online	0.3708	0.1464	-2.53	.0186*
UW Flexible Option	-0.2333	0.1054	-2.21	.0440*

*Note.* Results in this table represent combined results for two paired *t*-tests comparing each student’s average score on each dimension of Assessment A with their average score on each dimension of Assessment B. Therefore, a positive difference in the mean indicates a higher score on Assessment A than on Assessment B, and a negative difference in the mean signifies a higher score on Assessment B than on Assessment A. Due to differences in the way courses were structured, traditional online students submitted Assessment A before Assessment B. The majority of UW Flexible Option students, however, submitted Assessment B before Assessment A.

\*  $\alpha < .05$

### Discussion

The results of this study indicate that the students in the nonterm, direct assessment UW Flexible Option course demonstrated critical thinking at levels that are at least comparable to those demonstrated by students in a parallel traditional online course. If corroborated by additional research, these findings may help dispel concerns regarding the quality of CBE programs. The small sample size of the scored population and the restriction to only one course in a single CBE program means that these results should be replicated before they are assumed to be broadly applicable.

Furthermore, this study was not experimental in nature and made no effort to gauge changes in student ability. As a result, these findings do not demonstrate the efficacy of nonterm, direct assessment CBE as a learning environment. Rather, these findings merely demonstrate that upon course completion, the CBE students performed at a comparable or higher level than their traditional online counterparts. A variety of factors for which this study did not control could explain these results, such a difference in the academic or professional histories of the students, differing levels of student self-directedness or grit, or differences in advising or teaching support at any point in either version of the course.

Additionally, factors within this project itself complicate the interpretation of some of these results. Among these, the lack of demographic information on the traditional online student body rendered this study unable to control for the variety of student history variables that may otherwise

prove significant. However, other studies have found demographic indicators statistically insignificant in a similar population (Mayeshiba & Brower, 2017). The differential ordering of the assessments between the two versions of this course also complicates the interpretation of these results; however, because the traditional online students did not score higher on either assessment, this factor does not alter the essential findings.

In summary, while these findings should be corroborated, they do not support the idea that nonterm, direct assessment programs are categorically of lower quality when compared to more traditional programs. Indeed, these findings suggest that programs such as the UW Flexible Option that have deeply incorporated robust assessment strategies and high-quality student support may serve their students as well as or better than those in other teaching environments. For a previous generation of educators, investigations into new online learning environments demonstrated that what is now considered “traditional” online learning was not intrinsically better or worse than face-to-face instruction. Given these results, it may be that this is also the case for CBE and that eventually questions of quality will need to be rigorously addressed on a program-by-program basis, much as it is for other more traditional programs.

### **Acknowledgements**

The research presented here was supported by Lumina Foundation through their Grant for Research on the UW Flexible Option, Award #8423. The National Center for Research on Distance Education and Technological Advancements (DETA Center), established by the University of Wisconsin-Milwaukee (UWM), also supported this work and was supported by a grant from the U.S. Department of Education, Fund for the Improvement of Postsecondary Education (FIPSE), Award #P116Q140006. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies, their officers or employees.

## References

- Association of American Colleges and Universities. (2016). Valid assessment of learning in undergraduate education project. Retrieved from <https://www.aacu.org/value>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220. <http://dx.doi.org/10.1037/h0026256>
- Council for Adult and Experiential Learning (CAEL). (2015). *Competency-based education and quality assurance*. Chicago: CAEL. Retrieved from [http://cdn2.hubspot.net/hubfs/617695/CAEL\\_Reports/2014\\_CBE\\_Convening\\_Findings\\_Report.pdf](http://cdn2.hubspot.net/hubfs/617695/CAEL_Reports/2014_CBE_Convening_Findings_Report.pdf)
- Eduventures. (2014). *Closing the degree completion gap: Challenges and opportunities*. Boston, MA: Eduventures. Retrieved from <http://sparkroom.com/whitepapers/closing-degree-completion-gap-challenges-opportunities/>
- Fain, P. (2015, November 25). Measuring competency. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/news/2015/11/25/early-glimpse-student-achievement-college-america-competency-based-degree-provider>
- Gruppen, L., Burkhardt, J., Fitzgerald, J., Funnell, M., Haftel, H., Lypson, M., ... Vasquez, J. (2016). Competency-based education: Programme design and challenges to implementation. *Medical Education*, 50(5), 532–539. <https://doi.org/10.1111/medu.12977>
- Hart Research Associates. (2015). *Falling short? College learning and career success*. Retrieved from <https://www.luminafoundation.org/resources/falling-short>
- Johnstone, S., & Soares, L. (2014). Principles for developing competency-based education programs. *Change: The Magazine of Higher Learning*, 46(2), 12–19. doi: 10.1080/00091383.2014.896705
- Kelchen, R. (2015). *Competency-based education: Enrollments, demographics, and affordability*. Washington, DC: Center on Higher Education Reform, American Enterprise Institute. Retrieved from <http://hdl.voced.edu.au/10707/367090>
- Krause, J., Dias, L., & Schedler, C. (2016). A comparative study of competency-based courses demonstrating a potential measure of course quality and student success. *Distance Learning Administration 2016 Annual*, 117–125. Retrieved from <http://www.tiffanireardon.com/documents/2016-dl-proceedings-updated.pdf#page=117>
- LeBlanc, P., & Selbe, J. (2016, February 2). Another take on competency. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/views/2016/02/02/alternative-view-competency-based-education-essay>

- Lowry, R. (2016). Kappa as a measure of concordance in categorical sorting. *VassarStats*. Retrieved from <http://vassarstats.net/kappa.html>
- Mayeshiba, M., & Brower, A. (2017). Student success and retention using new definitions created for nonterm, direct assessment CBE. *Journal of Competency-Based Education*, 2(1). <https://doi.org/10.1002/cbe2.1039>
- Nodine, T. (2016). How did we get here? A brief history of competency-based higher education in the United States. *Journal of Competency-Based Education*, 1(1), 5–11. doi:10.1002/cbe2.1004
- State Higher Education Executive Officers Association. (2016). MSC demonstration year: September 2015 through August 2016. Retrieved from <http://www.sheeo.org/msc>
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360–363. Retrieved from <http://www.stfm.org/fmhub/fm2005/May/Anthony360.pdf>
- Ward, S. C. (2016, February 1). Let them eat cake (competently). *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/views/2016/02/01/competency-based-education-threatens-further-stratify-higher-education-essay>

